Chapter 7

Summary and conclusion

1 Viruses are obligate parasites that redirect the host cell machinery towards synthesis of their gene products and replication In order to overcome the restriction imposed for the use of eukaryotic translation system, viruses have evolved different strategies to express their genes Further, they have developed strategies that enable expression of maximum number of functional proteins from their limited genetic material The strategy of expression used in many of the plant viruses is not clearly understood In particular, the strategy of expression in the genus *Sobemoviruses* has not been elucidated

2 Capsid assembly is an important step in the life cycle of the virus Though structures of many icosahedral viruses have been determined at high resolution, the exact steps involved in the assembly process are not clearly understood

3 The present thesis deals with the mechanism of polyprotein processing and capsid assembly in *Sesbania mosaic virus* (SeMV) SeMV infects *Sesbania grandiflora* belonging to *Fabaceae* and is native to Andhra Pradesh, India. It is a single-stranded positive sense RNA virus with a genome length of 4149 nucleotides The genome encodes four potential overlapping open reading frames (ORFs) ORF1 codes for an 18 kDa protein that is proposed to be involved in the movement of the virus The ORF2 encodes a 105 kDa protein believed to be a polyprotein (single polypeptide chain having more than one functional protein) The ORF3, internal to ORF2 is proposed to be expressed as a *trans*-frame polyprotein, translation brought about by ribosomal frameshifting mechanism Coat protein (CP) is encoded by the fourth ORF present at the 3' end of the genome CP is responsible for the encapsidation and protection of the viral genome The capsids of SeMV are made of 180 copies of CP subunits that are built in an icosahedral geometry with T=3 symmetry

4 The objectives of the present study are as follows,

   a to elucidate the mechanism of polyprotein processing by SeMV protease domain through mutational analysis, and

b to delineate the relative contributions of metal-ion mediated interactions, N-terminal arginine-rich motif and the β-annulus in the assembly and stability of T=3 capsids

5 Although the strategy of polyprotein processing and the presence of protease domains have been predicted in many viruses based on sequence analysis, only in a few cases the functional role of the protease domain has been confirmed The processing of the ORF2 polyprotein of sobemoviruses is proposed to be catalyzed by the N-terminal serine protease domain However, there is no experimental evidence available to support this hypothesis In the present study the entire ORF2 was cloned and expressed in *E coli* It was observed that the polyprotein was cleaved into different domains by the protease domain The protease belonged to serine protease class with the active site catalytic triad residues, H181, D216 and S284 The role of these residues in catalysis was confirmed by site-directed mutagenesis

6 The amino acid sequence specificity of the protease cleavage was found to be at the carboxyl side of glutamic acid in E-T and E-S sites The protease cleaved the polyprotein at three different positions corresponding to E^{325}-T^{326}, E^{402}-T^{403} and E^{498}-S^{499}, thereby resulting in the generation of four proteins from a single polypeptide chain The various domains present in the ORF2 are arranged as follows, serine protease – viral protein genome linked (VPg) – 10 kDa protein (p10) – RNA dependent RNA polymerase (RdRP)

7 ORF2 encompasses an internal ORF, ORF3, which is proposed to be expressed by ribosomal frameshifting mechanism SeMV sequence contains a putative ribosomal frameshifting signal UUUAAAC followed by a stem loop structure, which would in principle lead to expression of another *trans*-frame polyprotein (ORF3) *In vitro* coupled transcription and translation experiments revealed that the ribosomal frameshift protein was indeed expressed Further to check if the *trans*-frame fusion protein could be proteolytically processed by the serine protease domain, it was expressed in *E coli* by incorporating an additional nucleotide to change the reading frame of ORF2 at the ribosomal

frameshift site The expressed *trans*-frame polyprotein (polyprotein-II) was also cleaved into individual domains by the N-terminal serine protease suggesting that *in planta*, two polyproteins that differ in their C-terminal domain might be expressed that upon processing by protease would result in different proteins The experiments presented in chapter 3 clearly demonstrate that protease is active in *cis* and is responsible for the proteolytic maturation of the polyproteins

8 Many of the viral proteases are also known to function in *trans* It was therefore of interest to examine the *trans*-catalytic activity of SeMV protease domain Expression of full-length SeMV protease domain in *E coli* resulted in the formation of soluble aggregates that were not amenable for purification Analysis of protease primary structure suggested that the N-terminal region of the protein was relatively hydrophobic and had the propensity to form transmembrane helix Hence, to check whether removal of the hydrophobic domain improves the solubility of protease, N-terminal 70 amino acid deletion mutant was generated (ΔN70P) In conformity with the prediction analysis, deletion of 70 amino acids from the N-terminus increased the solubility of the protease and the ΔN70P could be purified to homogeneity However, ΔN70P was unable to perform the *trans*-cleavage function Interestingly, the cleavage site mutant (E325A) of ΔN70PV fusion protein was active in *trans* Thus, the presence of VPg sequence along with the protease conferred the *trans*-catalytic activity on the protease

9 VPg is a small protein, which serves as protein primer for RNA synthesis in many animal and plant viruses. In SeMV, VPg is made of 77 amino acid residues An analysis of SeMV VPg sequence showed that it had a low hydrophobicity index and a large net charge (predicted pI value of 4 34), a property exhibited by "natively unfolded" proteins The biophysical studies on purified recombinant SeMV VPg showed that it lacked both secondary and tertiary structures ΔN70PV-E325A fusion protein showed a characteristic positive CD peak at 230 nm that was not observed in either ΔN70P or VPg

The conformational change was concomitant with change in the oligomeric status of the ΔN70P Mutational analysis suggested that stacking/ordering of aromatic residues between monomers of protease-VPg precursor might have resulted in the positive CD at 230 nm Enhanced *trans*-cleavage activity of the protease in fusion with the "natively unfolded" VPg might be a probable regulatory mechanism in the proteolytic maturation of polyprotein (chapter 4)

10 SeMV capsids are stabilized by RNA-protein, protein-protein and calcium mediated protein-protein interactions The removal of calcium has been proposed to be a prerequisite for the disassembly of the virus The crystal structure of native T=3 SeMV capsid revealed that residues D146 and D149 from one subunit and Y205, N267 and N268 of the neighboring subunit form the calcium-binding site (CBS) CBS environment is found to be identical even in the recombinant CP-NΔ65 T=1 capsids

11 The role of calcium and the residues involved in calcium co-ordination in the assembly and stability of T=3 and T=1 capsids was examined by mutational analysis Deletion of N267 and N268 did not affect both T=3 and T=1 assembly, although the capsids were devoid of calcium suggesting that assembly does not require calcium ions However, the stabilities of the capsids were reduced drastically Site-directed mutagenesis revealed that either a single mutation (D149N) or a double mutation (D146N-D149N) of SeMV CP drastically affected both the assembly and stability of T=3 capsids On the other hand, D146N-D149N mutation in CP-NΔ65 did not affect the assembly of T=1 capsids although their stability was reduced considerably Since the major difference between the T=3 and T=1 capsids is the absence of N-terminal arginine-rich motif (ARM) and β-annulus in the subunits forming the T=1 capsids, it is possible that D149 initiates the N-ARM - RNA interactions that lead to the formation of β-annulus essential for T=3 capsid assembly Detailed analyses of the calcium mutants are presented in chapter 5

12 The crystal structure of SeMV determined at 3 Å resolution revealed the three different symmetry related conformations exhibited by CP, A, B and C A

type of subunits make pentameric clusters at icosahedral five-fold, whereas, B and C subunits form hexamers at quasi-six-fold The CP can be divided into two domains, a random (R) domain comprising of first 65 residues and a shell (S) domain which forms the canonical eight-stranded anti parallel β-sheets commonly found in most viral CPs The R domain controls the size of the assembled capsids and it has two important motifs, ARM (residues 28-36) and amino acids residues responsible for the formation of β-annulus structure (residues 48-58) β-annulus is formed by three sets of β-interactions in which residues 48-52 from one C subunit, makes hydrogen-bonding interaction with residues 55-58 of the neighboring subunit The polypeptide backbone makes a 120° turn to enable the interaction possible between neighboring C-subunits, and a conserved proline residue (P53) makes a kink in the polypeptide fold The claim that β-annulus controls the assembly of T=3 capsids is solely based on the crystal structure data and there is no experimental evidence available till date to delineate the role of the amino acid residues that constitute the β-annulus structure in the assembly pathway

13 Mutational studies presented in chapter 6 show that ARM is responsible for RNA encapsidation and the mutation of all the arginine residues into glutamic acid residues results in the formation of less stable, empty T=3 capsids that are devoid of RNA Mutational studies of the residues involved in the formation of β-annulus structure showed that deletion of all the residues that make the β-annulus did not affect T=3 capsid assembly and the resulting VLPs were as stable as that of wild type VLPs Thus, the residues involved in the formation of β-annulus are dispensable for T=3 capsid assembly, and probably the ordering of the N-terminus occurs after the CP subunits assemble into capsids

14 Thus, in this study, the polyprotein processing mechanism was established for the first time in sobemoviruses The catalytic residues of serine protease and sites at which the protease cleaves the polyprotein were mapped and confirmed by mutational analysis By adopting the mechanism of polyprotein processing and ribosomal frameshifting, the virus could generate five different

proteins from ORF 2 and 3 The *trans*-catalytic activity of the protease was demonstrable only when it was expressed in fusion with the "natively-unfolded VPg" suggesting a role for VPg in the regulation of polyprotein processing With respect to the capsid assembly, the mutational analysis clearly demonstrated that calcium was not required for assembly but it played a predominant role in the stability of the capsids Another important finding of this study was that the residues involved in the formation of β-annulus structure were dispensable for both assembly and stability of T=3 capsids and ARM was responsible for the encapsidation of RNA but not required for T=3 capsid assembly *per se*